

A POSTERIORI CORRECTION OF FORECAST AND OBSERVATION ERROR VARIANCES

Leonid Rukhovets

*SAIC and Global Modeling and Assimilation Office, NASA/Goddard Space Flight Center,
Greenbelt, MD, USA*

1. Proposed method of total observation and forecast error variance correction is based on the assumption about normal distribution of "observed-minus-forecast" residuals (O-F), where O is an observed value and F is usually a short-term model forecast. This assumption can be accepted for several types of observations (except humidity) which are not grossly in error (Andersson and Järvinen 1999, Dharssi et al. 1992, Hollingsworth et al. 1986, Järvinen and Undén 1997, Lorenc and Hammon 1988).

Degree of nearness to normal distribution can be estimated by the *symmetry* or *skewness* (*luck of symmetry*) $a_3 = \mu_3/\sigma^3$

and *kurtosis* $a_4 = \mu_4/\sigma^4 - 3$

Here μ_i = i-order moment, σ is a standard deviation. It is well known that for normal distribution $a_3 = a_4 = 0$.

Table 1 contains a_3 and a_4 for O-F's of several types of observations: rawinsonde heights, winds and mixing ratio, aircraft winds, cloudtrack winds, and surface heights (recast as upper air) and winds. Six-hour model forecasts (F) were obtained using the Goddard Earth Observing System 4.0.3 assimilation run for October 1-31, 2003.

Figs.1-7 show O-F histograms for these observations (without gross errors).

Distributions of O-F's corresponding to rawinsonde heights and winds, aircraft winds, cloudtrack winds, surface heights and winds are close to the normal distribution. The rawinsonde mixing ratio distribution can not be considered normal. The *kurtosis* value for mixing ratio observations is also very large.

2. If a random variable X has normal distribution, then according to the statistical rules probability of X to be within the interval $(0, \alpha\sigma)$ is:

$$P(0 < X < \alpha\sigma) = F(\alpha) - F(0) = F(\alpha) - 0.5 \quad (1)$$

Here

$$F(x) = 1/\sqrt{2\pi} \int_{-\infty}^x e^{-t^2/2} dt,$$

σ - standard deviation of X, α - arbitrary constant, F - standard normal distribution function.

Suppose we have a number (percentage) of observations within some interval $(0, \alpha\sigma)$.

If the number does not correspond to (1), it can mean we are using σ that is not standard deviation for these observations.

However this number can be used to correct the standard deviations.

Table1. The *symmetry* a_3 and the *kurtosis* a_4 .

	rawinhght	rawinwind	rawinhumd	aircraftwind	cldtrkwind	srflheight	srflwind
a_3	0.13	-0.01	-0.08	0.01	-0.57	0.02	0.04
a_4	0.87	-1.12	6.08	-0.67	0.41	-1.07	-1.74

Corresponding author address:
Dr. Leonid Rukhovets, SAIC and
NASA/GSFC, Global Modeling and

Assimilation Office, Mail Code 610.1, MD
20771, USA

For this purpose it is convenient to apply so called "background check" procedure which is often used as a part of quality control in data assimilation systems (Andersson and Järvinen 1999, Dee et al. 1999).

The background check performs an examination of all observations against the short range (6 hours) forecasts. Actually, the inequality is verified for each observation:

$$(O - F)^2 < \alpha^2 \{(\tilde{\sigma}^o)^2 + (\tilde{\sigma}^f)^2\} \quad (2)$$

where α is a tolerance parameter, $\tilde{\sigma}^o$ and $\tilde{\sigma}^f$ are respectively appropriate prescribed observation and forecast error standard deviations. If for some observation the inequality (2) is not fulfilled, the observation is marked as suspect. The prescribed values $\tilde{\sigma}^o$ and $\tilde{\sigma}^f$ can approximately represent the true observation and forecast error standard deviations.

Let $\tilde{\sigma} = \{(\tilde{\sigma}^o)^2 + (\tilde{\sigma}^f)^2\}^{1/2}$ and $\sigma = \{(\sigma^o)^2 + (\sigma^f)^2\}^{1/2}$, where σ^o and σ^f are true standard deviations. Instead of (2) we can write:

$$|O - F| < \alpha(\tilde{\sigma}/\sigma)\sigma$$

Suppose M is the percentage of suspect observations which was obtained by some background check.

It means:

$$P(0 < O - F < \alpha\tilde{\sigma}) = 0.5 - M/(2*100)$$

But according to (1) we have:

$$P(0 < O - F < \alpha\tilde{\sigma}) = F(\alpha\tilde{\sigma}/\sigma) - 0.5$$

Then

$$F(\alpha\tilde{\sigma}/\sigma) = 1 - M/(2*100) \quad (3)$$

Using the table of standard normal distribution function we can find the value $\alpha\tilde{\sigma}/\sigma = m$, corresponding to $1 - M/(2*100)$.

$$\text{Then we can find } \sigma = \alpha\tilde{\sigma}/m \quad (4)$$

3. Consider one example. Let $\alpha = 2$. Suppose

a result of the background check gave suspect observation percentage of 4.6. Then $1 - a/(2*100) = 0.977$.

From (3) and the standard normal distribution table we have $2*\tilde{\sigma}/\sigma = 2$, and $\sigma = \tilde{\sigma}$. That is $\tilde{\sigma}$ is specified correctly. But it corresponds to the known statistical rule: about 4.6% of observations should be beyond 2σ .

4. Conclusions:

a) Using results of a background check the prescribed statistics of $(\tilde{\sigma}^o)^2 + (\tilde{\sigma}^f)^2$ can be corrected. Then the background check can be repeated with the new $\sigma = \{(\sigma^o)^2 + (\sigma^f)^2\}^{1/2}$.

b) The equation (4) together with the results of appropriate background check can be considered as a relation between observation and forecast error standard deviations. If the true value of σ^o is known, we can calculate σ^f .

c) Consider results of two background checks for the same observation type, but for different instruments. Then using (4) we can write two equations for true sigmas:

$$(\sigma_1^o)^2 + (\sigma_1^f)^2 = s_1$$

$$(\sigma_2^o)^2 + (\sigma_2^f)^2 = s_2$$

Because $\sigma_1^f = \sigma_2^f$ for both types of observations, subtraction of one equation from the other gives a relation between two observation error standard deviations. If one of them can be found easier than the other, the relation can be used to get the second one through the first one. For example, observation error standard deviation for TOVS heights can be found through the rawinsonde height observation error standard deviation. Analogously, cloud track

wind observation error standard deviation can be found through the rawinsonde wind observation error standard deviation.

REFERENCES:

- Andersson, E., and H. Järvinen, 1999: Variational quality control. *Q.J.R. Meteorol. Soc.*, 125, 697-722
- Dharssi, A. C., A. C. Lorenc, and N. B. Ingleby, 1992: Treatment of gross errors using maximum probability theory. *Q.J.R. Meteorol. Soc.*, 118, 1017-1036.
- Dee, D.P., L. Rukhovets, R. Todling, A.M. da Silva, and J.W. Larson, 1999: An adaptive buddy check for observational quality control. *Q.J.R. Meteorol. Soc.*, 127, 2451-2471
- Hollingsworth, A., D. B. Shaw, P. Lonnberg, L. Illari, K. Arpe and A. Simmons, 1986: Monitoring of observation and analysis quality by a data assimilation system. *Mon. Wea. Rev.*, 114, 861-879
- Järvinen, H., and P. Undén, 1997: Observation screening and first guess quality control in the ECMWF 3D-Var data assimilation system. ECMWF Tech. Memo. 236.
- Lorenc, A. C., and O. Hammon, 1988: Objective quality control of observations using Bayesian methods. Theory and practical implementation. *Q.J.R. Meteorol. Soc.* 114, 515-543

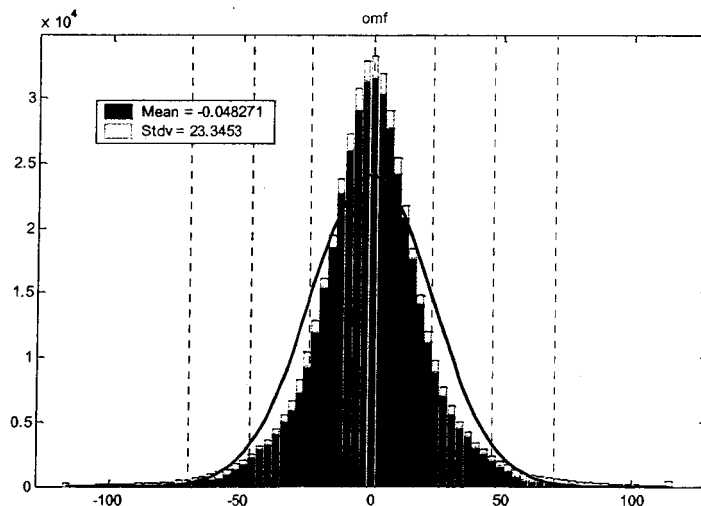


Figure 1. Histogram of O-F rawinsonde heights. Global. All levels. October 2003. Number of observations = 457919. Yellow color corresponds to observations that were marked as suspect by background check, but passed the adaptive buddy check (Dee et al. 1999). The blue curve shows the Gaussian distribution for the same mean and standard deviation.

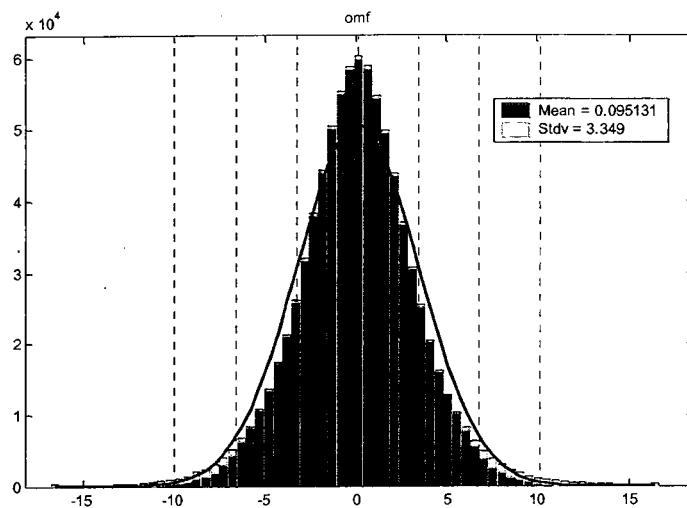


Figure 2. As in Fig.1 but for rawinsonde winds. Number of observations = 862146

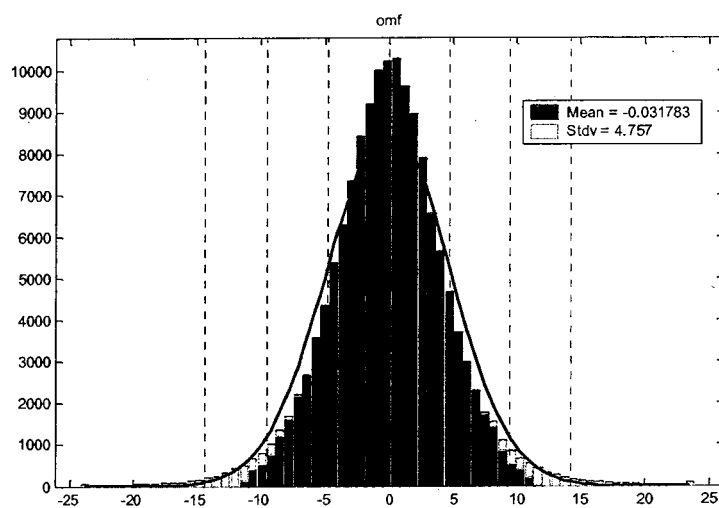


Figure 3. As in Fig.1 but for aircraft winds. Number of observations = 148612

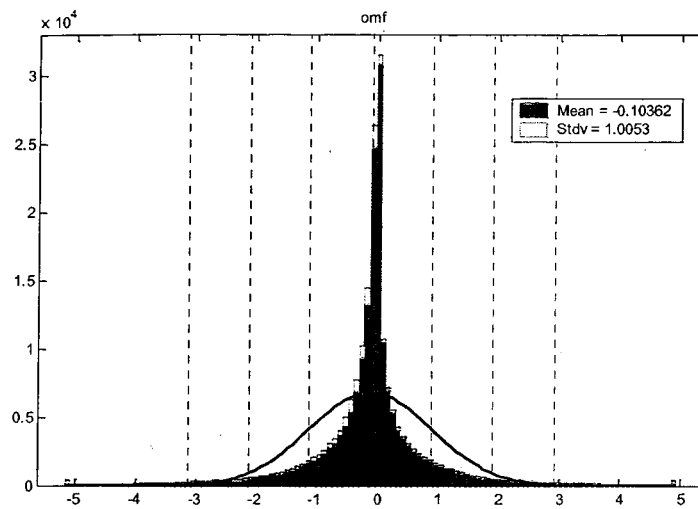


Figure 4. As in Fig.1 but for rawinsonde water vapor mixing ratio. Number of observations = 198624.

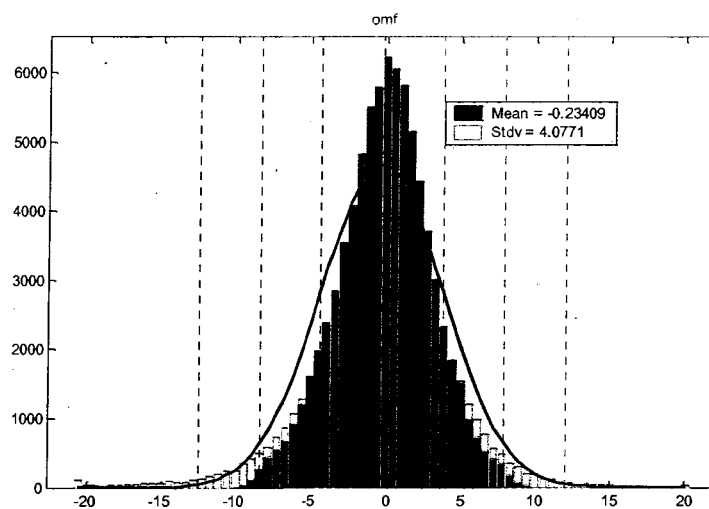


Figure 5. As in Fig.1 but for cloud track winds. Number of observations = 85776

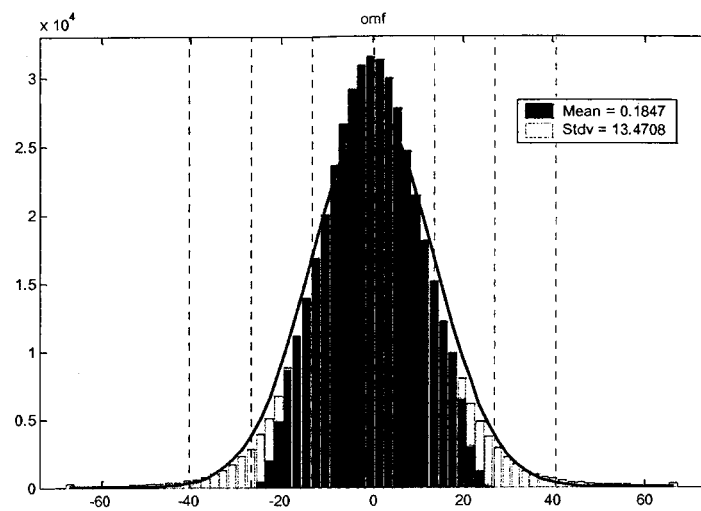


Figure 6. As in Fig.1 but for surface geopotential heights. Number of observations = 463369.

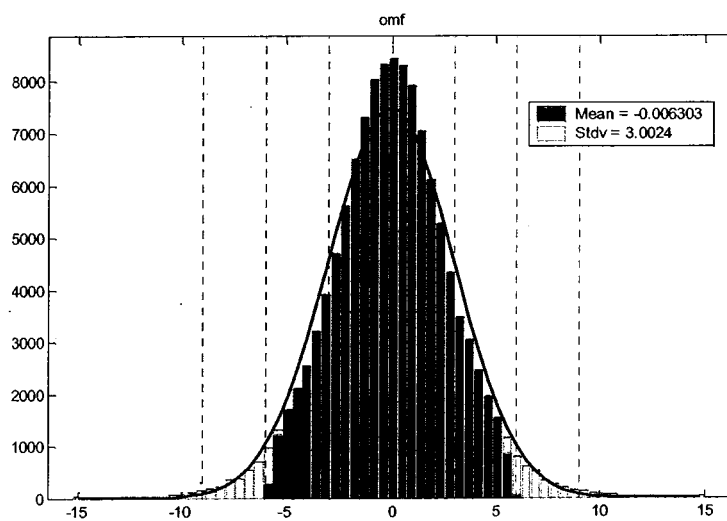


Figure 7. As in Fig.1 but for surface winds. Number of observations = 123254.